第四章 矩阵分解

第 13 讲 奇异值分解

黄定江

DaSE @ ECNU

djhuang@dase.ecnu.edu.cn

1 13.1 奇异值分解

2 13.2 基于奇异值分解的矩阵性质

3 13.3 奇异值分解与矩阵近似

13.1.1 奇异值分解的定义：历史

奇异值分解（Singular Value Decomposition，SVD）是线性代数和矩阵论中一种重要的矩阵分解技术。

- 1873 年，Beltrami 给出实正方阵的奇异值分解
- 1874 年，Jordan 也独立推导出实正方阵的奇异值分解
- 1902 年，Autonne 把奇异值分解推广到复方阵
- 1939 年，Eckhart 和 Young 进一步把它推广到复长方形矩阵

引入：图像压缩

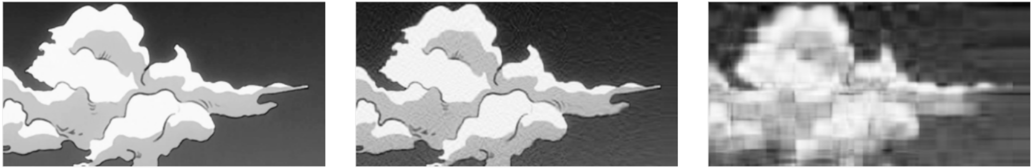
奇异值分解在数据分析、信号处理和模式识别等方面都具有广泛应用，比如在图像压缩领域：

- 图像数据中通常存在冗余，包括：图像中相邻像素间的相关性引起的空冗余；图像序列中不同帧之间存在相关性引起的时间冗余；不同彩色平面或频谱带的相关性引起的频谱冗余
- 可以通过图像压缩处理来减少图像数据中的冗余信息从而用更加高效的格式存储和传输数据，其原理就是通过图像矩阵分解理论减少表示数字图像时需要的数据量，比如通过矩阵的特征分解

引入：图像压缩

- 但是由于特征值分解压缩图片存在着不可靠性，所以通常会采用矩阵的奇异值分解，把获得的奇异值，取其中比较大的奇异值（类同特征值提取的压缩方法），舍去较小的奇异值，以达到数字图像压缩的目的
- 图像矩阵的奇异值及其特征空间反映了图像中的不同成分和特征。一般认为较大的奇异值及其对应的奇异向量表示图像信号，而噪声反映在较小的奇异值及其对应的奇异向量上
- 依据一定的准则选择门限，低于该门限的奇异值置零 (截断) ，然后通过这些奇异值和其对应的奇异向量重构图像进行去噪。若考虑图像的局部平稳性，也可以对图像分块进行奇异值分解去噪，这样能在一定程度上保护图像的边缘细节

引入：图像压缩



(a) 原图 (b) 提取 50 个奇异值的图像 (c) 提取 10 个奇异值的图像

图 1: 使用 SVD 进行图像压缩

定义 1

矩阵的奇异值分解是指，将一个非零的 $m \times n$ 实矩阵 $\boldsymbol{A}$，$\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 表示为以下三个实矩阵乘积的形式，即进行矩阵的因子分解：

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}} \tag{1}$$

其中 $\boldsymbol{U}$ 是 m 阶正交矩阵，$\boldsymbol{V}$ 是 n 阶正交矩阵，$\boldsymbol{\Sigma}$ 是由降序排列的非负的对角线元素组成的 $m \times n$ 矩形对角矩阵，满足

$$\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}} = \boldsymbol{I},\ \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{I},\ \boldsymbol{\Sigma} = diag(\sigma_1, \sigma_2, \cdots, \sigma_p),\ \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p,\ p = \min(m, n)$$

$\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ 称为矩阵 $\boldsymbol{A}$ 的奇异值分解，σ_i 称为 $\boldsymbol{A}$ 的奇异值，$\boldsymbol{U}$ 的列向量称为左奇异向量，$\boldsymbol{V}$ 的列向量称为右奇异向量。

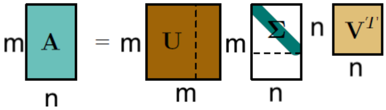


图 2: 完全奇异值分解

例 1

矩阵 $\boldsymbol{A}=\begin{pmatrix}1&0&0&0\\0&0&0&4\\0&3&0&0\\0&0&0&0\\2&0&0&0\end{pmatrix}$ 的奇异值分解为：

$$\begin{pmatrix}1&0&0&0\\0&0&0&4\\0&3&0&0\\0&0&0&0\\2&0&0&0\end{pmatrix}=\begin{pmatrix}0&0&\sqrt{0.2}&-\sqrt{0.8}&0\\1&0&0&0&0\\0&1&0&0&0\\0&0&0&0&1\\0&0&\sqrt{0.8}&\sqrt{0.2}&0\end{pmatrix}\begin{pmatrix}4&0&0&0\\0&3&0&0\\0&0&\sqrt{5}&0\\0&0&0&0\\0&0&0&0\end{pmatrix}\begin{pmatrix}0&0&0&1\\0&1&0&0\\1&0&0&0\\0&0&1&0\end{pmatrix}$$

13.1.2 奇异值分解的几何解释

从线性变换的角度理解奇异值分解，$m \times n$ 矩阵 $\boldsymbol{A}$ 表示从 n 维空间 $\mathbb{R}^n$ 到 m 维空间 $\mathbb{R}^m$ 的一个线性变换，

$$\mathcal{T}: \boldsymbol{x} \mapsto \boldsymbol{A}\boldsymbol{x}$$

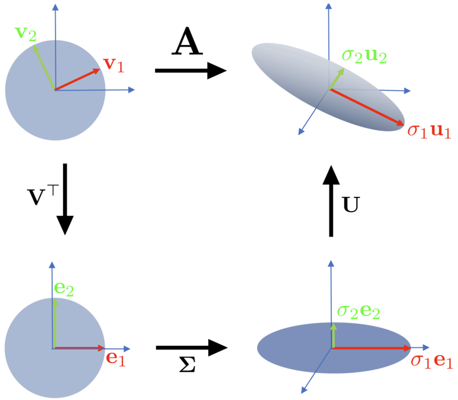
$\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^m$，$\boldsymbol{x}$ 和 $\boldsymbol{A}\boldsymbol{x}$ 分别是各自空间的向量。线性变换可以分解为三个简单的变换：

- 一个坐标系的旋转或者反射变换
- 一个坐标轴的缩放变换
- 另一个坐标系的旋转或者反射变换

奇异值定理保证这种分解一定存在。这就是奇异值分解的几何解释。

- 对矩阵 $\boldsymbol{A}$ 进行奇异值分解，得到 $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$， $\boldsymbol{V}$ 和 $\boldsymbol{U}$ 都是正交矩阵
- 所以 $\boldsymbol{V}$ 的列向量 $\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_n$ 构成 $\mathbb{R}^n$ 空间的一组标准正交基，表示 $\mathbb{R}^n$ 中的正交坐标系的旋转或反射变换
- $\boldsymbol{U}$ 的列向量 $\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_m$ 构成 $\mathbb{R}^m$ 空间的一组标准正交基，表示 $\mathbb{R}^m$ 中的正交坐标系的旋转或者反射变换
- $\boldsymbol{\Sigma}$ 的对角元素 $\sigma_1,\sigma_2,\ldots,\sigma_n$ 是一组非负实数，表示 $\mathbb{R}^n$ 中的原始正交坐标系坐标轴的 $\sigma_1,\sigma_2,\ldots,\sigma_n$ 倍的缩放变换。

- 任意一个向量 $\boldsymbol{x} \in \mathbb{R}^n$，经过基于 $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$ 的线性变换，等价于经过坐标系的旋转或者反射变换 $\boldsymbol{V}^{\mathrm{T}}$，坐标轴的缩放变换 $\boldsymbol{\Sigma}$，以及坐标系的旋转或者反射变换 $\boldsymbol{U}$ 得到向量 $\boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^m$。
- 对于 SVD 来说，分别改变了 $\mathbb{R}^n$ 和 $\mathbb{R}^m$ 两个空间的基底。而特征分解仅仅是在同一个空间中做变换。



下面通过一个例子直观地说明奇异值分解的几何意义。

例 2

给定一个 2 阶矩阵

$$\boldsymbol{A}=\begin{pmatrix}3 & 1\\ 2 & 1\end{pmatrix}$$

其奇异值分解为

$$\boldsymbol{U}=\begin{pmatrix}0.8174 & -0.5760\\ 0.5760 & 0.8174\end{pmatrix},\boldsymbol{\Sigma}=\begin{pmatrix}3.8643 & 0\\ 0 & 0.2588\end{pmatrix},\boldsymbol{V}=\begin{pmatrix}0.9327 & 0.3606\\ -0.3606 & 0.9327\end{pmatrix}$$

观察基于矩阵 $\boldsymbol{A}$ 的奇异值分解将 $\mathbb{R}^2$ 的标准正交基

$$\boldsymbol{e}_1=\begin{pmatrix}1\\ 0\end{pmatrix},\boldsymbol{e}_2=\begin{pmatrix}0\\ 1\end{pmatrix}$$

进行线性转换的情况。

首先，$\boldsymbol{V}^T$ 表示一个旋转变换，将标准正交基 $\boldsymbol{e}_1, \boldsymbol{e}_2$ 旋转，得到向量

$$\boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_1 = \begin{pmatrix} 0.9327 \\ -0.3606 \end{pmatrix}, \boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_2 = \begin{pmatrix} 0.3606 \\ 0.9327 \end{pmatrix}$$

其次，$\boldsymbol{\Sigma}$ 表示一个缩放变换，将向量 $\boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_1, \boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_2$ 在坐标轴方向缩放 σ_1 倍和 σ_2 倍，得到向量

$$\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_1 = \begin{pmatrix} 3.6042 \\ -0.0933 \end{pmatrix}, \boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_2 = \begin{pmatrix} 1.3935 \\ 0.2414 \end{pmatrix}$$

最后，$\boldsymbol{U}$ 表示一个旋转变换，再将向量 $\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_1, \boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_2$ 旋转得到

$$\boldsymbol{A}\boldsymbol{e}_1 = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_1 = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \boldsymbol{A}\boldsymbol{e}_2 = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{e}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

综上，矩阵的奇异值分解也可以看作是将其对应的线性变换分解为旋转变换、缩放变换以及旋转变换的组合。这一组合是一定存在的。

13.1.3 紧奇异值分解和截断奇异值分解

定义 2

设有 $m \times n$ 实矩阵 $\boldsymbol{A}$，其秩 $\operatorname{rank}(\boldsymbol{A}) = r, r \leq \min(m, n)$，则称 $\boldsymbol{U}_r \boldsymbol{\Sigma}_r \boldsymbol{V}_r^{\mathrm{T}}$ 为 $\boldsymbol{A}$ 的紧奇异值分解，即

$$\boldsymbol{A} = \boldsymbol{U}_r \boldsymbol{\Sigma}_r \boldsymbol{V}_r^{\mathrm{T}}$$

其中 $\boldsymbol{U}_r \in \mathbb{R}^{m \times r}, \boldsymbol{V}_r \in \mathbb{R}^{n \times r}$，$\boldsymbol{\Sigma}_r$ 是 r 阶对角矩阵；矩阵 $\boldsymbol{U}_r$ 由完全奇异值分解中 $\boldsymbol{U}$ 的前 r 列、矩阵 $\boldsymbol{V}_r$ 由 $\boldsymbol{V}$ 的前 r 列、矩阵 $\boldsymbol{\Sigma}_r$ 由 $\boldsymbol{\Sigma}$ 的前 r 个对角线元素得到。紧奇异值分解的对角矩阵 $\boldsymbol{\Sigma}_r$ 的秩与原始矩阵 $\boldsymbol{A}$ 的秩相等。

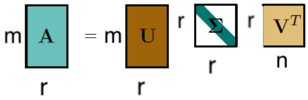


图 3: 紧 SVD

例 3

由例1给出的矩阵 $\boldsymbol{A}=\begin{pmatrix}1&0&0&0\\0&0&0&4\\0&3&0&0\\0&0&0&0\\2&0&0&0\end{pmatrix}$ 的秩 $r=3$，其紧奇异值分解为：

$$\begin{pmatrix}1&0&0&0\\0&0&0&4\\0&3&0&0\\0&0&0&0\\2&0&0&0\end{pmatrix}=\begin{pmatrix}0&0&\sqrt{0.2}\\1&0&0\\0&1&0\\0&0&0\\0&0&\sqrt{0.8}\end{pmatrix}\begin{pmatrix}4&0&0\\0&3&0\\0&0&\sqrt{5}\end{pmatrix}\begin{pmatrix}0&0&0&1\\0&1&0&0\\1&0&0&0\end{pmatrix}$$

定义 3

设有 $m \times n$ 实矩阵 $\boldsymbol{A}$，其秩 $\operatorname{rank}(\boldsymbol{A}) = r$，且 $0 < k < r$，则称 $\boldsymbol{U}_k \boldsymbol{\Sigma}_k \boldsymbol{V}_k^{\mathrm{T}}$ 为矩阵 $\boldsymbol{A}$ 的截断奇异值分解，即

$$\boldsymbol{A} \approx \boldsymbol{U}_k \boldsymbol{\Sigma}_k \boldsymbol{V}_k^{\mathrm{T}}$$

其中 $\boldsymbol{U}_k \in \mathbb{R}^{m \times k}$, $\boldsymbol{V}_k \in \mathbb{R}^{n \times k}$，$\boldsymbol{\Sigma}_k$ 是 k 阶对角矩阵；矩阵 $\boldsymbol{U}_k$ 由完全奇异值分解中 $\boldsymbol{U}$ 的前 k 列、矩阵 $\boldsymbol{V}_k$ 由 $\boldsymbol{V}$ 的前 k 列、矩阵 $\boldsymbol{\Sigma}_k$ 由 $\boldsymbol{\Sigma}$ 的前 k 个对角线元素得到。截断奇异值分解的对角矩阵 $\boldsymbol{\Sigma}_k$ 的秩比原始矩阵 $\boldsymbol{A}$ 的秩低。

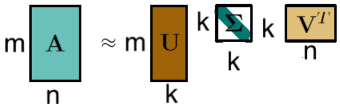


图 4: 截断 svd

例 4

由例1给出的矩阵 $\boldsymbol{A}=\begin{pmatrix}1&0&0&0\\0&0&0&4\\0&3&0&0\\0&0&0&0\\2&0&0&0\end{pmatrix}$ 的秩为 3，若取 $k=2$，则其截断奇异值分解为

$$\begin{pmatrix}1&0&0&0\\0&0&0&4\\0&3&0&0\\0&0&0&0\\2&0&0&0\end{pmatrix}\approx\begin{pmatrix}0&0\\1&0\\0&1\\0&0\\0&0\end{pmatrix}\begin{pmatrix}4&0\\0&3\end{pmatrix}\begin{pmatrix}0&0&0&1\\0&1&0&0\end{pmatrix}=\begin{pmatrix}0&0&0&0\\0&0&0&4\\0&3&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix}$$

在实际应用中，常常需要对矩阵的数据进行压缩，将其近似表示，奇异值分解提供了一种方法。后面将要叙述，奇异值分解是在平方损失意义下对矩阵的最优近似。紧奇异值对应着无损压缩，截断奇异值分解对应着有损压缩。

13.1.4 奇异值分解基本定理：对称矩阵的奇异值分解

对于对称矩阵来说，奇异值分解总是存在的。因为，我们知道如果 $\boldsymbol{A}$ 是对称矩阵，那么存在一个正交矩阵 $\boldsymbol{P}$ 使得 $\boldsymbol{A}$ 有特征分解

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Sigma}\boldsymbol{P}^T$$

此时我们令 $\boldsymbol{U} = \boldsymbol{P} = \boldsymbol{V}$ 那么就有

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$$

所以对称矩阵的奇异值分解就是他们的特征分解。
下面，我们将给出一般矩阵奇异值分解的存在性证明，并给出构造的一般方法。

奇异值分解基本定理

定理 1

奇异值分解基本定理 若 $\boldsymbol{A}$ 为一 $m \times n$ 实矩阵，$\boldsymbol{A} \in \mathbb{R}^{m \times n}$，则 $\boldsymbol{A}$ 的奇异值分解存在

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$$

其中 $\boldsymbol{U}$ 是 m 阶正交矩阵，$\boldsymbol{V}$ 是 n 阶正交矩阵，$\boldsymbol{\Sigma}$ 是 $m \times n$ 对角矩阵，其前 r 个对角元素 $(\sigma_1, \ldots, \sigma_r)$ 为正，且按降序排列，其余均为 0。

证明.

考虑矩阵 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$，这个矩阵是对称半正定的，所以我们可以对其进行谱分解

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}_n\boldsymbol{V}^{\mathrm{T}}$$

其中 $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ 是正交矩阵，$\boldsymbol{\Lambda}_n$ 是对称矩阵，并且对角线元素是 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的特征值 $\lambda_i \geq 0, i = 1, \ldots, n$，并且是按降序排列的。因为 $\mathrm{rank}(\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = r$，所以前 r 个特征值是正的。 □

注意到 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ 和 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 有相同的非零特征值，因此他们的秩是相等的。我们定义

$$\sigma_i = \sqrt{\lambda_i} > 0, i = 1, \ldots, r$$

记 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r$ 是 $\boldsymbol{V}$ 的前 r 列，它们同时也是 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 前 r 个特征值对应的特征向量。即有

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i, i = 1, \ldots, r$$

因此同时在两边左乘上 $\boldsymbol{A}$ 就有

$$(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{A}\boldsymbol{v}_i, i = 1, \ldots, r$$

这就意味着 $\boldsymbol{A}\boldsymbol{v}_i$ 是 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ 的特征向量，因为 $\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_j = \lambda_j \boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_j$ 所以这些特征向量也是正交的。所以将他们标准化则有

$$\boldsymbol{u}_i = \frac{\boldsymbol{A}\boldsymbol{v}_i}{\sqrt{\lambda_i}} = \frac{\boldsymbol{A}\boldsymbol{v}_i}{\sigma_i}, i = 1, \ldots, r$$

这些 $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r$ 是 r 个 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ 关于非零特征值 $\lambda_1, \ldots, \lambda_r$ 的特征向量。

因此

$$\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_j = \frac{1}{\sigma_i}\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_j = \frac{\lambda_j}{\sigma_i}\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_j = \begin{cases}\sigma_i & i=j\\ 0 & \text{otherwise}\end{cases}$$

以矩阵的方式重写即有

$$\begin{pmatrix}\boldsymbol{u}_1^{\mathrm{T}}\\ \vdots\\ \boldsymbol{u}_r^{\mathrm{T}}\end{pmatrix}\boldsymbol{A}\left(\boldsymbol{v}_1,\dots,\boldsymbol{v}_r\right) = diag(\sigma_1,\dots,\sigma_r) = \boldsymbol{\Sigma}_r \tag{2}$$

至此就证明了紧 SVD。我们下面继续证明完全 SVD。注意到根据定义

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_i = 0, i = r+1,\dots,n$$

即有

$$\boldsymbol{A}\boldsymbol{v}_i = 0, i = r+1,\dots,n$$

为了说明上述等式成立，我们假设 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_i=0$ 且 $\boldsymbol{A}\boldsymbol{v}_i\neq 0$，这意味着 $\boldsymbol{A}\boldsymbol{v}_i\in Null(\boldsymbol{A}^{\mathrm{T}})\equiv Col(\boldsymbol{A})^{\perp}$，这与 $\boldsymbol{A}\boldsymbol{v}_i\in Col(\boldsymbol{A})$ 矛盾。所以 $\boldsymbol{A}\boldsymbol{v}_i=0, i=r+1,\dots,n$。然后我们取相互正交的单位向量 $\boldsymbol{u}_{r+1},\dots,\boldsymbol{u}_m$ 均与 $\boldsymbol{u}_1,\dots,\boldsymbol{u}_r$ 正交，即有

$$\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_j=0, i=1,\dots,m; j=r+1,\dots,n$$

它们一起形成 $\mathbb{R}^m$ 的一组标准正交基。因此，扩展前述紧奇异值分解(2)有

$$\begin{pmatrix}\boldsymbol{u}_1^{\mathrm{T}}\\ \vdots\\ \boldsymbol{u}_m^{\mathrm{T}}\end{pmatrix}\boldsymbol{A}\left(\boldsymbol{v}_1,\dots,\boldsymbol{v}_n\right)=\begin{pmatrix}\boldsymbol{\Sigma}_r & \boldsymbol{0}^{\mathrm{T}}\\ \boldsymbol{0} & \boldsymbol{O}\end{pmatrix}=\boldsymbol{\Sigma}$$

令 $\boldsymbol{U}=\left(\boldsymbol{u}_1,\dots,\boldsymbol{u}_m\right), \boldsymbol{V}=\left(\boldsymbol{v}_1,\dots,\boldsymbol{v}_n\right)$ 就能得到 SVD 分解

$$\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$$

至此就证明了矩阵 $\boldsymbol{A}$ 存在奇异值分解。

13.1.5 奇异值分解的计算

奇异值分解定理的证明过程蕴含了奇异值分解的计算方法。矩阵 $\boldsymbol{A}$ 的奇异值分解可以通过求对称矩阵 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的特征值和特征向量得到。$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的特征向量构成正交矩阵 $\boldsymbol{V}$ 的列；$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的特征值 λ_j 的平方根为奇异值 σ_j，即

$$\sigma_j = \sqrt{\lambda_j}, j = 1, 2, \cdots, n$$

对其由大到小排列作为对角线元素，构成对角矩阵 $\boldsymbol{\Sigma}$；求正奇异值对应的左奇异向量，再求扩充的 $\boldsymbol{A}^{\mathrm{T}}$ 的标准正交基，构成正交矩阵 $\boldsymbol{U}$ 的列。从而得到 $\boldsymbol{A}$ 的奇异值分解 $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$。

给定 $m \times n$ 矩阵 $\boldsymbol{A}$，可以根据上面的叙述写出奇异值分解的计算过程

- 首先求 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的特征值和特征向量。计算对称矩阵 $\boldsymbol{W}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$，求解特征方程
$$(\boldsymbol{W}-\lambda \boldsymbol{I})\boldsymbol{x}=0$$
得到特征值 λ_i，并将特征值由大到小排列
$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$
将特征值 $\lambda_i(i=1,2,\ldots,n)$ 代入特征方程求得对应的特征向量。
- 求 n 阶正交矩阵 $\boldsymbol{V}$。将特征向量单位化得到 $\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_n$，构成 n 阶正交矩阵 $\boldsymbol{V}$ 即
$$\boldsymbol{V}=\begin{pmatrix}\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_n\end{pmatrix}$$
- 求 $m \times n$ 对角矩阵 $\boldsymbol{\Sigma}$。计算 $\boldsymbol{A}$ 的奇异值
$$\sigma_i=\sqrt{\lambda_i}, i=1,2,\ldots,n$$
构造 $m \times n$ 矩形对角矩阵 $\boldsymbol{\Sigma}$，主对角线元素是奇异值，其余元素是零
$$\boldsymbol{\Sigma}=diag(\sigma_1,\sigma_2,\ldots,\sigma_n)$$

- 求 m 阶正交矩阵 $\boldsymbol{U}$, 对 $\boldsymbol{A}$ 的前 r 个正奇异值，令

$$\boldsymbol{u}_j = \frac{1}{\sigma_j}\boldsymbol{A}\boldsymbol{v}_j, j = 1, 2, \ldots, r$$

得到

$$\boldsymbol{U}_1 = (\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r)$$

求 $\boldsymbol{A}^{\mathrm{T}}$ 的零空间的一组标准正交基 $\{\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_m\}$，令

$$\boldsymbol{U}_2 = (\boldsymbol{u}_{r+1}, \ldots, \boldsymbol{u}_m)$$

并令

$$\boldsymbol{U} = (\boldsymbol{U}_1, \boldsymbol{U}_2)$$

- 得到奇异值分解

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$$

例 5

试求矩阵 $\boldsymbol{A}=\begin{pmatrix}1 & 1\\ 2 & 2\\ 0 & 0\end{pmatrix}$ 的奇异值分解

解

求对称矩阵

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\begin{pmatrix}5 & 5\\ 5 & 5\end{pmatrix}$$

求 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的特征值与特征向量，即求

$$\lambda^2-10\lambda=0$$

所以特征值为 $\lambda_1=10,\lambda_2=0$ 从而得到 $\boldsymbol{v}_1=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ 1\end{pmatrix},\boldsymbol{v}_2=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ -1\end{pmatrix}$ 所以正交矩阵

$$\boldsymbol{V}=\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix},$$

奇异值为 $\sigma_1 = \sqrt{10}, \sigma_2 = 0$ 所以对角矩阵为

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sqrt{10} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

再求正交矩阵 $\boldsymbol{U}$，基于 $\boldsymbol{A}$ 的正奇异值计算得到列向量

$$\boldsymbol{u}_1 = \frac{1}{\sigma_1}\boldsymbol{A}\boldsymbol{v}_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$$

而列向量 $\boldsymbol{u}_2, \boldsymbol{u}_3$ 是 $\boldsymbol{A}^{\mathrm{T}}$ 零空间 $\mathsf{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right)$ 的一组标准正交基，所以

$$\boldsymbol{u}_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \boldsymbol{u}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

故正交矩阵 $\boldsymbol{U}$ 为

$$\boldsymbol{U}=\frac{1}{\sqrt{5}}\begin{pmatrix}1 & -2 & 0\\ 2 & 1 & 0\\ 0 & 0 & \sqrt{5}\end{pmatrix}$$

所以 $\boldsymbol{A}$ 的奇异值分解为

$$\boldsymbol{A}=\frac{1}{\sqrt{5}}\begin{pmatrix}1 & -2 & 0\\ 2 & 1 & 0\\ 0 & 0 & \sqrt{5}\end{pmatrix}\begin{pmatrix}\sqrt{10} & 0\\ 0 & 0\\ 0 & 0\end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}$$

13.1.6 奇异值分解和特征分解：几个性质

性质 1

设矩阵 $\boldsymbol{A}$ 的奇异值分解为 $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$，则以下关系成立：

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}})=\boldsymbol{V}(\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{\Sigma})\boldsymbol{V}^{\mathrm{T}}$$

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}=(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}})(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}})^{\mathrm{T}}=\boldsymbol{U}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{\mathrm{T}})\boldsymbol{U}^{\mathrm{T}}$$

也就是说，矩阵 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}, \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ 的特征分解存在，且可以由矩阵 $\boldsymbol{A}$ 的奇异值分解的矩阵表示。$\boldsymbol{V}$ 的列向量是 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的特征向量，$\boldsymbol{U}$ 的列向量是 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ 的特征向量，$\boldsymbol{\Sigma}$ 是奇异值是 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}, \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ 的特征值的平方根。

性质 2

矩阵 $\boldsymbol{A}$ 的奇异值分解中，奇异值 $\sigma_1, \sigma_2, \ldots, \sigma_n$ 是唯一的，而矩阵 $\boldsymbol{U}, \boldsymbol{V}$ 不是唯一的。

在矩阵 $\boldsymbol{A}$ 的奇异值分解中，奇异值、左奇异向量和右奇异向量之间存在对应关系。

性质 3

设矩阵 $\boldsymbol{A}$ 的奇异值分解为 $\boldsymbol{A}=\boldsymbol{U\Sigma V}^{\mathrm{T}}$，则以下关系成立：

$$\boldsymbol{A}\boldsymbol{v}_j=\sigma_j\boldsymbol{u}_j, j=1,2,\ldots,n$$

$$\begin{cases}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}_j=\sigma_j\boldsymbol{v}_j & j=1,2,\ldots,n\\ \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}_j=0 & j=n+1,n+2,\ldots,m\end{cases}$$

证明.

由 $\boldsymbol{A}=\boldsymbol{U\Sigma V}^{\mathrm{T}}$，易知 $\boldsymbol{AV}=\boldsymbol{U\Sigma}$。比较这一等式两端的第 j 列，得到 $\boldsymbol{A}\boldsymbol{v}_j=\sigma_j\boldsymbol{u}_j, j=1,2,\ldots,n$，这是矩阵 $\boldsymbol{A}$ 的右奇异向量和奇异值、左奇异向量的关系。类似地，我们可以得到另外一组关于矩阵 $\boldsymbol{A}$ 的左奇异向量和奇异值、右奇异向量的关系 □

奇异值分解和特征分解的关系

考虑矩阵 $\boldsymbol{A}$ 的特征分解 $\boldsymbol{A}=\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$ 和奇异值分解 $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$。

- 对于任何矩阵 $\boldsymbol{A}\in\mathbb{R}^{n\times m}$，SVD 始终存在。特征分解仅针对方阵 $\boldsymbol{A}\in\mathbb{R}^{n\times n}$ 定义的，并且只有在我们可以找到 n 个相互独立的特征向量时才存在。
- 特征分解矩阵中的向量不一定是正交的，因此对基的改变并不是简单的旋转和缩放。另一方面，SVD 中矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 是正交矩阵，因此它们可以表示旋转或反射。
- 特征分解和 SVD 都是三个线性映射的组合：

1. 改变空间的基底
2. 在每个新基底方向上进行独立缩放并且从一个空间映射到另外一个空间。
3. 改变另外一个空间的基底

- 特征分解和 SVD 之间的主要区别在于，在 SVD 中，上述两个空间可以是不同维的向量空间。

- 在 SVD 中，左右奇异向量矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 通常不是互为逆矩阵。在特征分解中，特征向量矩阵 $\boldsymbol{P}$ 和 $\boldsymbol{P}^{-1}$ 是互为逆矩阵。
- 在 SVD 中，对角矩阵 $\boldsymbol{\Sigma}$ 中的项都是实数且非负，对于特征分解中的对角矩阵来说通常不成立。
- SVD 和特征分解通过他们的投影被紧密联系
  - $\boldsymbol{A}$ 的左奇异向量是 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ 的特征向量
  - $\boldsymbol{A}$ 的右奇异向量是 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的特征向量
  - $\boldsymbol{A}$ 非零奇异值是 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 非零特征值的开方，同时也是 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ 非零特征值的开方
- 对于对称矩阵的特征分解和 SVD 是相同的。

本小节，我们将利用矩阵 $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ 的完全奇异值分解

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$$

或紧奇异值分解

$$\boldsymbol{A} = \boldsymbol{U}_r\boldsymbol{\Sigma}_r\boldsymbol{V}_r{}^T$$

来重新探讨关于矩阵 $\boldsymbol{A}$ 的一些性质：

- 矩阵 $\boldsymbol{A}$ 的秩、零空间和列空间
- 矩阵范数
- 矩阵广义逆
- 正交投影

13.2.1 秩、零空间和列空间

性质 4

设矩阵 $\boldsymbol{A} \in \mathbb{R}^{m \times n}$，其奇异值分解为 $\boldsymbol{A} = \boldsymbol{U\Sigma V}^T$，则矩阵 $\boldsymbol{A}$ 的秩和对角矩阵 $\boldsymbol{\Sigma}$ 的秩相等，等于正奇异值 σ_i 的个数 r（包含重复的奇异值）。

同样，由于在实际中 $\boldsymbol{\Sigma}$ 上对角元可能很小，但不为零（例如由于数值误差），因此可以在给定的误差 $\epsilon \geq 0$ 的范围内给出一个更加可靠的数值秩：

$$r = \max_{\sigma_k > \epsilon\sigma_1} k$$

性质 5

设矩阵 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ 的紧奇异值分解为 $\boldsymbol{A}=\boldsymbol{U}_r\boldsymbol{\Sigma}_r\boldsymbol{V}_r^T$，其秩 $\mathrm{rank}(\boldsymbol{A})=r$，则有 $Null(\boldsymbol{A})=n-r$ 且生成 $Null(\boldsymbol{A})$ 的一组正交基底由 $\boldsymbol{V}$ 的最后 $n-r$ 列给出，也即

$$Null(\boldsymbol{A})=Col(\boldsymbol{V}_{nr}),\quad \boldsymbol{V}_{nr}=(\boldsymbol{v}_{r+1},\dots,\boldsymbol{v}_n)$$

证明.

根据线性代数的基本定理，有 $\mathrm{Null}(\boldsymbol{A})=n-r$。因为 $V=(\boldsymbol{V}_r\,\boldsymbol{V}_{nr})$ 是正交矩阵，所以 $\{\boldsymbol{v}_{r+1},\dots,\boldsymbol{v}_n\}$ 是正交向量组，并且 $\boldsymbol{V}_r^{\mathrm{T}}\boldsymbol{V}_{nr}=0$。因此对于 $\boldsymbol{V}_{nr}$ 列空间中任意的向量 $\eta=\boldsymbol{V}_{nr}z$，由矩阵 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ 的紧奇异值分解有

$$\boldsymbol{A}\eta=\boldsymbol{U}_r\boldsymbol{\Sigma}_r\boldsymbol{V}_r^{\mathrm{T}}\eta=\boldsymbol{U}_r\boldsymbol{\Sigma}_r\boldsymbol{V}_r^{\mathrm{T}}\boldsymbol{V}_{nr}z=0$$

所以

$$\mathrm{Null}(\boldsymbol{A})=\mathrm{Col}(\boldsymbol{V}_{nr})$$

□

性质 6

设矩阵 $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ 的紧奇异值分解为 $\boldsymbol{A}=\boldsymbol{U}_r\boldsymbol{\Sigma}_r\boldsymbol{V}_r^T$，其秩 $\operatorname{rank}(\boldsymbol{A})=r$，则 $\boldsymbol{A}$ 的列空间由 $\boldsymbol{U}$ 的前 r 个列向量生成，即

$$\mathit{Col}\,(\boldsymbol{A})=\mathit{Col}\,(\boldsymbol{U}_r)\,,\,\boldsymbol{U}_r=(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_r)$$

证明.

首先，因为 $\boldsymbol{\Sigma}_r\boldsymbol{V}_r^{\mathrm{T}}\in\mathbb{R}^{r\times n}$ $r\le n$，是一个行满秩矩阵，则当 $\boldsymbol{x}$ 张成整个 $\mathbb{R}^n$ 时，$\boldsymbol{z}=\boldsymbol{\Sigma}_r\boldsymbol{V}_r^{\mathrm{T}}\boldsymbol{x}$ 张成整个 $\mathbb{R}^r$。因此

$$\begin{aligned}\mathrm{Col}\,(\boldsymbol{A})&=\{\boldsymbol{y}|\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x},\boldsymbol{x}\in\mathbb{R}^n\}\\&=\{\boldsymbol{y}|\boldsymbol{y}=\boldsymbol{U}_r\boldsymbol{\Sigma}_r\boldsymbol{V}_r^{\mathrm{T}}\boldsymbol{x},\boldsymbol{x}\in\mathbb{R}^n\}\\&=\{\boldsymbol{y}|\boldsymbol{y}=\boldsymbol{U}_r\boldsymbol{z},\boldsymbol{z}\in\mathbb{R}^r\}\\&=\mathrm{Col}\,(\boldsymbol{U}_r)\end{aligned}$$

□

例 6

矩阵 $\boldsymbol{A}=\begin{pmatrix}1&0&0&0\\0&0&0&4\\0&3&0&0\\0&0&0&0\\2&0&0&0\end{pmatrix}$ 的奇异值分解为

$$\begin{pmatrix}1&0&0&0\\0&0&0&4\\0&3&0&0\\0&0&0&0\\2&0&0&0\end{pmatrix}=\begin{pmatrix}0&0&\sqrt{0.2}&-\sqrt{0.8}&0\\1&0&0&0&0\\0&1&0&0&0\\0&0&0&0&1\\0&0&\sqrt{0.8}&\sqrt{0.2}&0\end{pmatrix}\begin{pmatrix}4&0&0&0\\0&3&0&0\\0&0&\sqrt{5}&0\\0&0&0&0\\0&0&0&0\end{pmatrix}\begin{pmatrix}0&0&0&1\\0&1&0&0\\1&0&0&0\\0&0&1&0\end{pmatrix}$$

所以 $\mathsf{Col}\,(\boldsymbol{A}) = \mathsf{Col}\begin{pmatrix} 0 & 0 & \sqrt{0.2} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sqrt{0.8} \end{pmatrix}, \mathsf{Null}\,(\boldsymbol{A}) = \mathsf{Col}\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$

13.2.2 矩阵范数

- 矩阵的 F 范数满足以下等式

$$\|\boldsymbol{A}\|_F^2 = \mathrm{Tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \sum_{i=1}^{n} \sigma_i^2$$

其中 σ_i 是矩阵 $\boldsymbol{A}$ 的奇异值。因此 F 范数的平方实际上就是奇异值的平方和。

- 矩阵 2 范数的平方是 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 的最大特征值，所以

$$\|\boldsymbol{A}\|_2^2 = \sigma_1^2$$

即 $\boldsymbol{A}$ 的 2 范数就是 $\boldsymbol{A}$ 的最大的奇异值

- 对于矩阵核范数，我们有

$$\|\boldsymbol{A}\|_* = \sum_{i=1}^{r} \sigma_i, r = \mathrm{rank}(\boldsymbol{A})$$

核范数常常出现在低秩矩阵补全和秩最小化问题中，这是因为 $\|\boldsymbol{A}\|_*$ 表示在谱范数以 1 为界的矩阵集合中 $\mathrm{rank}(\boldsymbol{A})$ 的最大可能凸下界。

13.2.3 广义逆

定义 4

令 $\boldsymbol{A}$ 是一个 $m \times n$ 矩阵，若存在一个的 $n \times m$ 矩阵 $\boldsymbol{G}$，使得下列条件满足：

$$(\boldsymbol{A}\boldsymbol{G})^T = \boldsymbol{A}\boldsymbol{G}$$

$$(\boldsymbol{G}\boldsymbol{A})^T = \boldsymbol{G}\boldsymbol{A}$$

$$\boldsymbol{G}\boldsymbol{A}\boldsymbol{G} = \boldsymbol{G}$$

$$\boldsymbol{A}\boldsymbol{G}\boldsymbol{A} = \boldsymbol{A}$$

则称 $\boldsymbol{G}$ 是 $\boldsymbol{A}$ 的广义逆或 *Moore-Penrose* 逆或伪逆。

我们还可以定义其他广义逆，比如在上面四条中去掉一条到三条就可以定义另外 14 种广义逆。但是只有 Moore-Penrose 逆有下列性质。

性质 7

设矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，如果 $\boldsymbol{G}$ 是 $\boldsymbol{A}$ 的 *Moore-Penrose* 逆，那么 $\boldsymbol{G}$ 是 $\boldsymbol{A}$ 唯一的 *Moore-Penrose* 逆。

在后面的内容中，我们不关心其他的广义逆，所以默认这里的广义逆均指的是 Moore-Penrose 逆。

利用奇异值分解求解广义逆

若矩阵 $\boldsymbol{M}$ 的奇异值分解为 $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$，那么 $\boldsymbol{M}$ 的伪逆为

$$\boldsymbol{M}^\dagger = \boldsymbol{V}\boldsymbol{\Sigma}^\dagger\boldsymbol{U}^T$$

其中 $\boldsymbol{\Sigma}^\dagger$ 是 $\boldsymbol{\Sigma}$ 的伪逆，是将 $\boldsymbol{\Sigma}$ 主对角线上每个非零元素都求倒数之后再转置得到的。

例 7

求矩阵 $\boldsymbol{A}=\begin{bmatrix}0 & 1\\ 1 & 1\\ 1 & 0\end{bmatrix}$ 的广义逆

解

首先我们对 $\boldsymbol{A}$ 进行奇异值分解得

$$\boldsymbol{A}=\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T=\begin{bmatrix}\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}}\\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}}\\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}}\end{bmatrix}\begin{bmatrix}\sqrt{3} & 0\\ 0 & 1\\ 0 & 0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}$$

根据公式我们有

$$\boldsymbol{A}^{\dagger}=\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{3}} & 0 & 0\\ 0 & 1 & 0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}}\\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}}\end{bmatrix}=\frac{1}{3}\begin{bmatrix}-1 & 1 & 2\\ 2 & 1 & -1\end{bmatrix}$$

对于一些特殊的矩阵的逆，我们可以使用奇异值分解推导出更便于计算的公式。

性质 8

如果 $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 可逆，那么

$$\boldsymbol{A}^{\dagger} = \boldsymbol{A}^{-1}$$

例 8

矩阵 $\boldsymbol{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ 的广义逆矩阵为

$$\boldsymbol{A}^{\dagger} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{3} \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

容易验证

$$\boldsymbol{A}\boldsymbol{A}^{\dagger} = \boldsymbol{A}^{\dagger}\boldsymbol{A} = \boldsymbol{I}$$

性质 9

当矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 为列满秩矩阵时有

$$\boldsymbol{A}^{\dagger} = (\boldsymbol{A}^{T}\boldsymbol{A})^{-1}\boldsymbol{A}^{T}$$

证明.

如果 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 是一个列满秩矩阵，因此 $r = n \leq m$，故有

$$\boldsymbol{A}^{\dagger}\boldsymbol{A} = \boldsymbol{V}_r\boldsymbol{V}_r^{T} = \boldsymbol{I}_n$$

所以 $\boldsymbol{A}^{\dagger}$ 是矩阵 $\boldsymbol{A}$ 的左逆（即 $\boldsymbol{A}^{\dagger}\boldsymbol{A} = \boldsymbol{I}_n$）。注意到 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 是可逆的，所以

$$(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{T} = (\boldsymbol{V}\boldsymbol{\Sigma}^{-2}\boldsymbol{V}^{T})\boldsymbol{V}\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{T}} = \boldsymbol{A}^{\dagger}$$

□

$\boldsymbol{A}$ 所有的左逆都可以表示为 $\boldsymbol{A}^{li} = \boldsymbol{A}^{\dagger} + \boldsymbol{Q}^{T}$ 其中 $\boldsymbol{Q}$ 满足 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q} = 0$

性质 10

当矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 为行满秩矩阵时有

$$\boldsymbol{A}^{\dagger} = \boldsymbol{A}^{T}(\boldsymbol{A}\boldsymbol{A}^{T})^{-1}$$

证明.

如果 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 是一个行满秩矩阵，因此 $r = m \leq n$，故有

$$\boldsymbol{A}\boldsymbol{A}^{\dagger} = \boldsymbol{U}_r\boldsymbol{U}_r^{T} = \boldsymbol{I}_m$$

所以 $\boldsymbol{A}^{\dagger}$ 是矩阵 $\boldsymbol{A}$ 的右逆（即 $\boldsymbol{A}\boldsymbol{A}^{\dagger} = \boldsymbol{I}_m$）。注意到 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ 是可逆的，所以

$$\boldsymbol{A}^{T}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{-1} = \boldsymbol{V}\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}(\boldsymbol{U}\boldsymbol{\Sigma}^{-2}\boldsymbol{U}^{T}) = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{\mathrm{T}} = \boldsymbol{A}^{\dagger}$$

□

$\boldsymbol{A}$ 所有的右逆都可以表示为 $\boldsymbol{A}^{ri} = \boldsymbol{A}^{\dagger} + \boldsymbol{Q}$ 其中 $\boldsymbol{Q}$ 满足 $\boldsymbol{A}\boldsymbol{Q} = 0$

例 9

求矩阵 $\boldsymbol{A}=\begin{bmatrix}0 & 1\\1 & 1\\1 & 0\end{bmatrix}$ 的广义逆

解

显然这是一个列满秩的矩阵，我们利用列满秩矩阵的公式

$$\boldsymbol{A}^{\dagger}=(\boldsymbol{A}^{T}\boldsymbol{A})^{-1}\boldsymbol{A}^{T}=(\begin{bmatrix}0 & 1 & 1\\1 & 1 & 0\end{bmatrix}\begin{bmatrix}0 & 1\\1 & 1\\1 & 0\end{bmatrix})^{-1}\begin{bmatrix}0 & 1 & 1\\1 & 1 & 0\end{bmatrix}$$

$$=\begin{bmatrix}1 & 2\\2 & 1\end{bmatrix}^{-1}\begin{bmatrix}0 & 1 & 1\\1 & 1 & 0\end{bmatrix}$$

$$=\frac{1}{3}\begin{bmatrix}-1 & 1 & 2\\2 & 1 & -1\end{bmatrix}$$

13.2.4 正交投影

我们知道任何一个矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 是一个从输入空间 $\mathbb{R}^n$ 到输出空间 $\mathbb{R}^m$ 的线性映射，并且根据线性代数基本定理，我们可以将 $\mathbb{R}^n, \mathbb{R}^m$ 分解成如下正交子空间的直和：

$$\mathbb{R}^n = \mathsf{Null}\left(\boldsymbol{A}\right) \oplus \mathsf{Null}\left(\boldsymbol{A}\right)^{\perp} = \mathsf{Null}\left(\boldsymbol{A}\right) \oplus \mathsf{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right)$$

$$\mathbb{R}^m = \mathsf{Col}\left(\boldsymbol{A}\right) \oplus \mathsf{Col}\left(\boldsymbol{A}\right)^{\perp} = \mathsf{Col}\left(\boldsymbol{A}\right) \oplus \mathsf{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right)$$

正如前面讨论的，矩阵 $\boldsymbol{A}$ 的奇异值分解 $\boldsymbol{A} = \boldsymbol{U\Sigma V}^T$ 给四个基本子空间提供了正交基底，我们令

$$\boldsymbol{U} = (\boldsymbol{U}_r, \boldsymbol{U}_{nr}), \boldsymbol{V} = (\boldsymbol{V}_r, \boldsymbol{V}_{nr})$$

其中 $r = \mathrm{rank}(\boldsymbol{A})$，我们就有

$$\mathsf{Null}\left(\boldsymbol{A}\right) = \mathsf{Col}\left(\boldsymbol{V}_{nr}\right), \mathsf{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right) = \mathsf{Col}\left(\boldsymbol{V}_r\right)$$

$$\mathsf{Col}\left(\boldsymbol{A}\right) = \mathsf{Col}\left(\boldsymbol{U}_r\right), \mathsf{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right) = \mathsf{Col}\left(\boldsymbol{U}_{nr}\right)$$

接下来，我们讨论如何将一个向量 $\boldsymbol{x} \in \mathbb{R}^n$ 投影到 $\mathsf{Null}\left(\boldsymbol{A}\right), \mathsf{Col}\left(\boldsymbol{A}^T\right)$ 中，以及把一个向量 $\boldsymbol{y} \in \mathbb{R}^m$ 投影到 $\mathsf{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right), \mathsf{Col}\left(\boldsymbol{A}\right)$ 中。

将向量投影到子空间

如果给定一个向量 $\boldsymbol{x} \in \mathbb{R}^n$ 和 d 个线性无关的向量 $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d \in \mathbb{R}^n$，那么 $\boldsymbol{x}$ 到子空间 $span\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d\}$ 的正交投影就是向量

$$\boldsymbol{x}^* = \boldsymbol{B}\boldsymbol{\alpha}$$

其中 $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d), \boldsymbol{\alpha} \in \mathbb{R}^d$, 并且我们需要解方程

$$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{\alpha} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{x}$$

来得到 $\boldsymbol{\alpha}$。注意到，如果 $\boldsymbol{B}$ 的列向量是正交的，则有 $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} = \boldsymbol{I}_d$，因此

$$\boldsymbol{\alpha} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{x}$$

故可得投影

$$\boldsymbol{x}^* = \boldsymbol{B}\boldsymbol{B}^T\boldsymbol{x}$$

将向量 $\boldsymbol{x} \in \mathbb{R}^n$ 投影到 Null $(\boldsymbol{A})$ 上

因此如果我们将一个向量 $\boldsymbol{x} \in \mathbb{R}^n$ 投影到 Null $(\boldsymbol{A})$ 上，投影 $\pi_{\mathsf{Null}(\boldsymbol{A})}(\boldsymbol{x})$ 则可以通过以下等式算出

$$\pi_{\mathsf{Null}(\boldsymbol{A})}(\boldsymbol{x}) = (\boldsymbol{V}_{nr}\boldsymbol{V}_{nr}^{\mathrm{T}})\boldsymbol{x}$$

我们又知道

$$\boldsymbol{I} = \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{V}_r\boldsymbol{V}_r^{\mathrm{T}} + \boldsymbol{V}_{nr}\boldsymbol{V}_{nr}^{\mathrm{T}}$$

因此由广义逆的定义，我们可知投影矩阵

$$\boldsymbol{P}_{\mathsf{Null}(\boldsymbol{A})} = (\boldsymbol{V}_{nr}\boldsymbol{V}_{nr}^{\mathrm{T}}) = \boldsymbol{I}_n - \boldsymbol{V}_r\boldsymbol{V}_r^{\mathrm{T}} = \boldsymbol{I}_n - \boldsymbol{A}^{\dagger}\boldsymbol{A}$$

在 $\boldsymbol{A}$ 行满秩的情况下，由性质 10 可知 $\boldsymbol{A}^{\dagger} = \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{-1}$，所以有

$$\boldsymbol{P}_{\mathsf{Null}(\boldsymbol{A})} = \boldsymbol{I}_n - \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{-1}\boldsymbol{A}$$

矩阵 $\boldsymbol{P}_{\mathsf{Null}(\boldsymbol{A})}$ 称为子空间 Null $(\boldsymbol{A})$ 上的正交投影。

将向量 $\boldsymbol{x} \in \mathbb{R}^n$ 投影到 $\mathsf{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right)$ 上

用同样的方式，我们可以得到 $\boldsymbol{x}$ 在 $\mathsf{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right)$ 上的投影为 $\pi_{\mathsf{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right)}(\boldsymbol{x}) = \left(\boldsymbol{V}_r \boldsymbol{V}_r^{\mathrm{T}}\right)\boldsymbol{x} = \boldsymbol{A}^{\dagger}\boldsymbol{A}\boldsymbol{x}$

而当 $\boldsymbol{A}$ 行满秩的时，有 $\pi_{\mathsf{Col}\left(\boldsymbol{A}^{\mathrm{T}}\right)}(\boldsymbol{x}) = \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{-1}\boldsymbol{A}\boldsymbol{x}$

将向量 $\boldsymbol{y} \in \mathbb{R}^m$ 投影到 $\mathsf{Col}\left(\boldsymbol{A}\right)$ 上

类似的，我们可以得到 $\boldsymbol{y}$ 在 $\mathsf{Col}\left(\boldsymbol{A}\right)$ 上的投影为 $\pi_{\mathsf{Col}(\boldsymbol{A})}(\boldsymbol{y}) = \left(\boldsymbol{U}_r \boldsymbol{U}_r^{\mathrm{T}}\right)\boldsymbol{y} = \boldsymbol{A}\boldsymbol{A}^{\dagger}\boldsymbol{y}$

而当 $\boldsymbol{A}$ 列满秩的时，有 $\pi_{\mathsf{Col}(\boldsymbol{A})}(\boldsymbol{y}) = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$

将向量 $\boldsymbol{y} \in \mathbb{R}^m$ 投影到 $\mathsf{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right)$ 上

向量 $\boldsymbol{y}$ 在 $\mathsf{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right)$ 上的投影为 $\pi_{\mathsf{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right)}(\boldsymbol{y}) = \left(\boldsymbol{U}_{nr} \boldsymbol{U}_{nr}^{\mathrm{T}}\right)\boldsymbol{y} = (\boldsymbol{I}_m - \boldsymbol{A}\boldsymbol{A}^{\dagger})\boldsymbol{y}$

并且当 $\boldsymbol{A}$ 列满秩时，有 $\pi_{\mathsf{Null}\left(\boldsymbol{A}^{\mathrm{T}}\right)}(\boldsymbol{y}) = (\boldsymbol{I}_m - \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}})\boldsymbol{y}$

13.3.1 奇异值分解与低秩表示

考虑通过奇异值分解把矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ 表示为一些更低秩矩阵的和。

- 假设矩阵 $\boldsymbol{A}$ 的奇异值分解为 $\boldsymbol{A} = \boldsymbol{U\Sigma V}^T$，其中 $\boldsymbol{U} \in \mathbb{R}^{m\times n}$ 和 $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ 都是正交矩阵，$\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ 是对角矩阵。我们把 $\boldsymbol{A}$ 的奇异值分解看成矩阵 $\boldsymbol{U\Sigma}$ 和 $\boldsymbol{V}^{\mathrm{T}}$ 的乘积，将 $\boldsymbol{U\Sigma}$ 按列分块，将 $\boldsymbol{V}^{\mathrm{T}}$ 按行分块，即

$$\boldsymbol{U\Sigma} = (\sigma_1 \boldsymbol{u}_1, \sigma_2 \boldsymbol{u}_2, \ldots, \sigma_n \boldsymbol{u}_n);\ \boldsymbol{V}^{\mathrm{T}} = \begin{pmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \boldsymbol{v}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_n^{\mathrm{T}} \end{pmatrix}$$

则有

$$\boldsymbol{A} = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \cdots + \sigma_n \boldsymbol{u}_n \boldsymbol{v}_n^T = \sum_{i=1}^{n} \sigma_i \boldsymbol{A}_i$$

称该式子为矩阵 $\boldsymbol{A}$ 的外积展开式，其中 $\boldsymbol{A}_i = \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{T}}$ 为 $m \times n$ 的秩 1 矩阵，是列向量 $\boldsymbol{u}_i$ 和行向量 $\boldsymbol{v}_i^{\mathrm{T}}$ 的外积，其第 k 行第 j 列元素为 $\boldsymbol{u}_i$ 的第 k 个元素与 $\boldsymbol{v}_i^{\mathrm{T}}$ 的第 j 个元素的乘积。

- 如果矩阵 $\boldsymbol{A}$ 的秩为 r，则对于任意 $i > r$ 的项，因为奇异值为 0，所以可以将该矩阵分解为 r 个秩为 1 矩阵 $\boldsymbol{A}_i$ 之和：
$$\boldsymbol{A} = \sum_{i=1}^{r} \sigma_i \boldsymbol{A}_i$$
其中外积矩阵 $\boldsymbol{A}_i$ 前面的系数是矩阵 $\boldsymbol{A}$ 第 i 个非零奇异值 σ_i。
- 更进一步，如果把上述 $\boldsymbol{A}_i$ 从 1 到 r 求和替换成从 1 到 $k(k < r)$ 求和，则我们可以获得矩阵 $\boldsymbol{A}$ 的近似
$$\hat{\boldsymbol{A}}(k) = \sum_{i=1}^{k} \sigma_i \boldsymbol{A}_i$$
其中 $\text{rank}(\hat{\boldsymbol{A}}) = k$，称为矩阵 $\boldsymbol{A}$ 的秩 k 近似。

问题

如何度量矩阵 $\boldsymbol{A}$ 和它的秩 k 近似 $\hat{\boldsymbol{A}}(k)$ 之间的差异大小或近似程度？能否找到矩阵 $\boldsymbol{A}$ 的一个最优低秩矩阵近似？

13.3.2 低秩矩阵近似

低秩矩阵近似问题

给定一个秩为 r 的矩阵 $\boldsymbol{A}$，欲求其最优的秩 k 近似矩阵 $\hat{\boldsymbol{A}}(k)$，其中 $k \leq r$，该问题可形式化为求

$$\min_{\hat{\boldsymbol{A}}(k) \in \boldsymbol{R}^{m,n}} \|\boldsymbol{A} - \hat{\boldsymbol{A}}(k)\|_F,$$

$$\text{s.t.:} \quad \operatorname{rank}(\hat{\boldsymbol{A}}(k)) = k$$

定理 2

设矩阵 $\boldsymbol{A} \in \mathbb{R}^{n \times n}$，矩阵的秩 $\operatorname{rank}(\boldsymbol{A}) = r$，并设 $\mathbb{M}$ 为 $\mathbb{R}^{m \times n}$ 中所有秩不超过 k 的矩阵集合 $0 < k < r$，则存在一个秩为 k 的矩阵 $\boldsymbol{X} \in \mathbb{M}$，使得

$$\|\boldsymbol{A} - \boldsymbol{X}\|_F = \min_{\boldsymbol{S} \in \mathbb{M}} \|\boldsymbol{A} - \boldsymbol{S}\|_F$$

称矩阵 $\boldsymbol{X}$ 为矩阵 $\boldsymbol{A}$ 在 F 范数下的最优近似。

定理 3

设矩阵 $\boldsymbol{A} \in \mathbb{R}^{m\times n}$，矩阵的秩 $\operatorname{rank}(\boldsymbol{A}) = r$，有奇异值分解 $\boldsymbol{A} = \boldsymbol{U\Sigma V}^T$，并设 $\mathbb{M}$ 为 $\mathbb{R}^{m\times n}$ 中所有秩不超过 k 的矩阵集合 $0 < k < r$，若秩为 k 的矩阵 $\boldsymbol{X} \in \mathbb{M}$，满足

$$\|\boldsymbol{A} - \boldsymbol{X}\|_F = \min_{\boldsymbol{S}\in\mathbb{M}} \|\boldsymbol{A} - \boldsymbol{S}\|_F$$

则 $\|\boldsymbol{A} - \boldsymbol{X}\|_F = (\sigma_{k+1}^2 + \sigma_{k+2}^2 + \cdots + \sigma_n^2)^{\frac{1}{2}}$。特别地，若 $\boldsymbol{A}' = \boldsymbol{U\Sigma}'\boldsymbol{V}^T$，其中

$$\boldsymbol{\Sigma}' = \begin{pmatrix} \sigma_1 & & & & & \\ & \ddots & & & 0 & \\ & & \sigma_k & & & \\ & & & 0 & & \\ & 0 & & & \ddots & \\ & & & & & 0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_k & 0 \\ 0 & 0 \end{pmatrix}$$

则 $\|\boldsymbol{A} - \boldsymbol{A}'\|_F = (\sigma_{k+1}^2 + \sigma_{k+2}^2 + \cdots + \sigma_n^2)^{\frac{1}{2}} = \min_{\boldsymbol{S}\in\mathbb{M}} \|\boldsymbol{A} - \boldsymbol{S}\|_F$。

证明.

若秩为 k 的矩阵 $\boldsymbol{X} \in \mathbb{M}$, 满足 $\|\boldsymbol{A}-\boldsymbol{X}\|_F = \min_{\boldsymbol{S}\in\mathbb{M}} \|\boldsymbol{A}-\boldsymbol{S}\|_F$ 则

$$\|\boldsymbol{A}-\boldsymbol{X}\|_F \leq \|\boldsymbol{A}-\boldsymbol{A}'\|_F = (\sigma_{k+1}^2+\sigma_{k+2}^2+\cdots+\sigma_n^2)^{\frac{1}{2}}$$

下面证明 $\|\boldsymbol{A}-\boldsymbol{X}\|_F \geq (\sigma_{k+1}^2+\sigma_{k+2}^2+\cdots+\sigma_n^2)^{\frac{1}{2}}$。设 $\boldsymbol{X}$ 的奇异值分解为 $\boldsymbol{Q\Omega P}^T$，其中

$$\boldsymbol{\Omega} = \begin{pmatrix} \omega_1 & & & & & \\ & \ddots & & & 0 & \\ & & \omega_k & & & \\ & & & 0 & & \\ & 0 & & & \ddots & \\ & & & & & 0 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Omega}_k & 0 \\ 0 & 0 \end{pmatrix}$$

□

若令矩阵 $\boldsymbol{B}=\boldsymbol{Q}^T\boldsymbol{A}\boldsymbol{P}$，则 $\boldsymbol{A}=\boldsymbol{Q}\boldsymbol{B}\boldsymbol{P}^T$，由此得到 $\|\boldsymbol{A}-\boldsymbol{X}\|_F=\|\boldsymbol{Q}(\boldsymbol{B}-\boldsymbol{\Omega})\boldsymbol{P}^T\|_F=\|\boldsymbol{B}-\boldsymbol{\Omega}\|_F$。用 $\boldsymbol{\Omega}$ 分块方法对 $\boldsymbol{B}$ 分块

$$\boldsymbol{B}=\begin{pmatrix}\boldsymbol{B}_{11} & \boldsymbol{B}_{12}\\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22}\end{pmatrix}$$

其中 $\boldsymbol{B}_{11}\in\mathbb{R}^{k\times k},\boldsymbol{B}_{12}\in\mathbb{R}^{k\times(n-k)},\boldsymbol{B}_{21}\in\mathbb{R}^{(n-k)\times k},\boldsymbol{B}_{22}\in\mathbb{R}^{(n-k)\times(n-k)}$ 可得

$$\|\boldsymbol{A}-\boldsymbol{X}\|_F^2=\|\boldsymbol{B}-\boldsymbol{\Omega}\|_F^2=\|\boldsymbol{B}_{11}-\boldsymbol{\Omega}_k\|_F^2+\|\boldsymbol{B}_{12}\|_F^2+\|\boldsymbol{B}_{21}\|_F^2+\|\boldsymbol{B}_{22}\|_F^2$$

现证 $\boldsymbol{B}_{12}=0,\boldsymbol{B}_{21}=0$ 用反证法。若 $\boldsymbol{B}_{12}\neq 0$，令

$$\boldsymbol{Y}=\boldsymbol{Q}\begin{pmatrix}\boldsymbol{B}_{11} & \boldsymbol{B}_{12}\\ 0 & 0\end{pmatrix}\boldsymbol{P}^T$$

则 $\boldsymbol{Y}\in\mathbb{M}$ 且 $\|\boldsymbol{A}-\boldsymbol{Y}\|_F^2=\|\boldsymbol{B}_{21}\|_F^2+\|\boldsymbol{B}_{22}\|_F^2<\|\boldsymbol{A}-\boldsymbol{X}\|_F^2$，这与 $\boldsymbol{X}$ 的定义 $\|\boldsymbol{A}-\boldsymbol{X}\|_F=\min_{\boldsymbol{S}\in\mathbb{M}}\|\boldsymbol{A}-\boldsymbol{S}\|_F$ 矛盾，因此 $\boldsymbol{B}_{12}=0$。

同理可证 $\boldsymbol{B}_{21}=0$。于是

$$\|\boldsymbol{A}-\boldsymbol{X}\|_F^2=\|\boldsymbol{B}_{11}-\boldsymbol{\Omega}_k\|_F^2+\|\boldsymbol{B}_{22}\|_F^2$$

再证 $\boldsymbol{B}_{11}=\boldsymbol{\Omega}_k$，为此令

$$\boldsymbol{Z}=\boldsymbol{Q}\begin{pmatrix}\boldsymbol{B}_{11} & 0\\ 0 & 0\end{pmatrix}\boldsymbol{P}^T$$

则 $\boldsymbol{Z}\in\mathbb{M}$，且

$$\|\boldsymbol{A}-\boldsymbol{Z}\|_F^2=\|\boldsymbol{B}_{22}\|_F^2\leq\|\boldsymbol{B}_{11}-\boldsymbol{\Omega}_k\|_F^2+\|\boldsymbol{B}_{22}\|_F^2=\|\boldsymbol{A}-\boldsymbol{X}\|_F^2$$

由 $\boldsymbol{X}$ 的定义

$$\|\boldsymbol{A}-\boldsymbol{X}\|_F=\min_{\boldsymbol{S}\in\mathbb{M}}\|\boldsymbol{A}-\boldsymbol{S}\|_F$$

知

$$\|\boldsymbol{B}_{11}-\boldsymbol{\Omega}_k\|_F^2=0$$

即 $\boldsymbol{B}_{11}=\boldsymbol{\Omega}_k$。

最后看 $\boldsymbol{B}_{22}$，设 $\boldsymbol{B}_{22}$ 有奇异值分解为 $\boldsymbol{U}_1\boldsymbol{\Lambda}\boldsymbol{V}_1^T$，则 $\|\boldsymbol{A}-\boldsymbol{X}\|_F=\|\boldsymbol{B}_{22}\|_F=\|\boldsymbol{\Lambda}\|_F$。下面证明 $\boldsymbol{\Lambda}$ 的对角线元素为 $\boldsymbol{A}$ 的奇异值。为此令

$$\boldsymbol{U}_2=\begin{pmatrix}\boldsymbol{I}_k & 0\\ 0 & \boldsymbol{U}_1\end{pmatrix},\boldsymbol{V}_2=\begin{pmatrix}\boldsymbol{I}_k & 0\\ 0 & \boldsymbol{V}_1\end{pmatrix}$$

其中 $\boldsymbol{I}_k$ 是 k 阶单位矩阵，$\boldsymbol{U}_2,\boldsymbol{V}_2$ 的分块与 $\boldsymbol{B}$ 的分块一致。注意到 $\boldsymbol{B}$ 以及 $\boldsymbol{B}_{22}$ 的奇异值分解，即得

$$\boldsymbol{U}_2^T\boldsymbol{Q}^T\boldsymbol{A}\boldsymbol{P}\boldsymbol{V}_2=\begin{pmatrix}\boldsymbol{\Omega}_k & \\ & \boldsymbol{\Lambda}\end{pmatrix},\boldsymbol{A}=(\boldsymbol{Q}\boldsymbol{U}_2)\begin{pmatrix}\boldsymbol{\Omega}_k & \\ & \boldsymbol{\Lambda}\end{pmatrix}(\boldsymbol{P}\boldsymbol{V}_2)^T$$

由此可知 $\boldsymbol{\Lambda}$ 的对角线元素为 $\boldsymbol{A}$ 的奇异值故有

$$\|\boldsymbol{A}-\boldsymbol{X}\|_F=\|\boldsymbol{\Lambda}\|_F\geq(\sigma_{k+1}^2+\sigma_{k+2}^2+\cdots+\sigma_n^2)^{\frac{1}{2}}$$

可证 $\|\boldsymbol{A}-\boldsymbol{X}\|_F=(\sigma_{k+1}^2+\sigma_{k+2}^2+\cdots+\sigma_n^2)^{\frac{1}{2}}=\|\boldsymbol{A}-\boldsymbol{A}'\|_F$

- 在秩不超过 k 的 $m \times n$ 矩阵的集合中，存在矩阵 $\boldsymbol{A}$ 的 F 范数意义下的最优近似矩阵 $\boldsymbol{X}$
- $\boldsymbol{A}' = \boldsymbol{U}\boldsymbol{\Sigma}'\boldsymbol{V}^T$ 是达到最优值的一个矩阵
- 紧奇异值分解是在 F 范数意义下的无损压缩
- 截断奇异值分解是有损压缩
- 截断奇异值分解得到的矩阵的秩为 k，通常远小于原始矩阵的秩 r，所以是由低秩矩阵实现了对原始矩阵的压缩。

注记

定理3中若把 F 范数改为谱范数，则有

$$\|\boldsymbol{A} - \boldsymbol{X}\|_2 = \sigma_{k+1} = \min_{\boldsymbol{S} \in \mathbb{M}} \|\boldsymbol{A} - \boldsymbol{S}\|_2,$$

成立。定理3也被称为 Eckhart-Young 或 Eckhart-Young-Mirsky 定理。

例 10

求例1中矩阵 $\boldsymbol{A}=\begin{pmatrix}1&0&0&0\\0&0&0&4\\0&3&0&0\\0&0&0&0\\2&0&0&0\end{pmatrix}$ 秩为 2 的最优近似。

解

先对 $\boldsymbol{A}$ 进行奇异值分解得

$$\begin{pmatrix} 0 & 0 & -\frac{\sqrt{5}}{5} & 0 & -\frac{2\sqrt{5}}{5} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -\frac{2\sqrt{5}}{5} & 0 & \frac{\sqrt{5}}{5} \end{pmatrix} \begin{pmatrix} 4 & & & \\ & 3 & & \\ & & \sqrt{5} & \\ & & & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{pmatrix}$$

然后令

$$\boldsymbol{A}' = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

即为矩阵 $\boldsymbol{A}$ 秩 2 的最优近似。

基于奇异值分解的图像压缩

- 假定一幅图像有 $m \times n$ 个像素，如果将这 mn 个数据一起传送，往往会显得数据量太大。因此，我们希望能够改为传送另外一些比较少的数据，并且在接收端还能够利用这些传送的数据重构原图像。
- 用 $m \times n$ 矩阵 $\boldsymbol{A}$ 表示要传送的原 $m \times n$ 个像素。假定对矩阵 $\boldsymbol{A}$ 进行奇异值分解，便得到 $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$，其中，奇异值按照从大到小的顺序排列。如果从中选择 k 个大奇异值以及与这些奇异值对应的左和右奇异向量逼近原图像，便可以共使用 $k(n+m+1)$ 个数值代替原来的 $m \times n$ 个图像数据。这 $k(n+m+1)$ 个被选择的新数据是矩阵 $\boldsymbol{A}$ 的前 k 个奇异值、$m \times m$ 左奇异向量矩阵 $\boldsymbol{U}$ 的前 k 列和 $n \times n$ 右奇异向量矩阵 $\boldsymbol{V}$ 的前 k 列的元素。
- 把比率

$$\rho = \frac{nm}{k(n+m+1)} \tag{3}$$

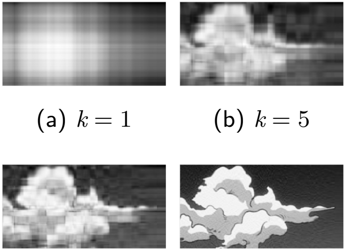
称为图像的压缩比。显然，被选择的大奇异值的个数 k 应该满足条件 $k(n+m+1) < nm$，即 $k < \frac{nm}{n+m+1}$。

例 11

右边四张图在视觉上展示了取不同数量的奇异值的效果：

- 当 $k=5$ 时，我们已经可以看出图像上是什么了。
- 当 $k=10$ 时，我们获得了更多的细节。但是仍然有一些模糊。
- 当 $k=50$ 时，我们获得了一个相当不错的图像，只有非常细微的地方有一些模糊。整体上和原图相差无几。

原图是一张 1328×680 的图像，要传输这样一张图像需要发送 $1328\times 680=903040$ 个数值。而如果使用 $k=50$ 时的截断 *SVD*，那么只需要传送 $50\times(1328+680+1)=100450$ 个数值。也就是说压缩比达到了 8.9899。



(a) $k=1$　(b) $k=5$

(c) $k=10$　(d) $k=50$

图 5: 不同 k 值对于压缩图像的影响

- 因此，我们在传送图像的过程中，就无须传送 $m \times n$ 个原始数据，而只需要传送 $k(n+m+1)$ 个有关奇异值和奇异向量的数据即可。在接收端，在接收到奇异值 $\sigma_1, \sigma_2, \cdots, \sigma_k$ 以及左奇异向量 $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_k$ 和右奇异向量 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_k$ 后，即可通过截尾的奇异值分解公式

$$\hat{\boldsymbol{A}} = \sum_{i=1}^{k} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{T}} \tag{4}$$

  重构出原图像。
- 一个容易理解的事实是：若 k 值偏小，即压缩比 p 偏大，则重构的图像的质量有可能不能令人满意。反之，过大的 k 值又会导致压缩比过小，从而降低图像压缩和传送的效率。因此，需要根据不同种类的图像，选择合适的压缩比，以兼顾图像传送效率和重构质量。

本讲小结

分解的形式

- 完全奇异值分解
- 紧奇异值分解
- 截断奇异值分解
- $\cdots$

与奇异值分解相关的矩阵性质

- 矩阵范数
- 广义逆
- 最优低秩矩阵近似
- $\cdots$

奇异值分解将用于矩阵的计算问题：线性方程组求解、最小二乘问题等！

奇异值分解与优化：矩阵的低秩近似问题实际上是一个优化问题。它表明数据科学与机器学习中某些优化问题可以方便的通过奇异值分解进行求解，这些优化问题还包括 PCA、正交的 Procrustean 变换等！